Міністерство освіти і науки України
Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського"
Фізико-технічний інститут

“КРИПТОГРАФІЯ”

Комп’ютерний практикум
Робота № 1

Виконав:

студент групи ФБ-93 Проценко О.А.

Перевірила:

Селюх П. В.

Київ – 2021

1. **Назва роботи**

   Експериментальна оцінка ентропії на символ джерела відкритого тексту.

   Мета
   Засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.
2. Особливості роботи:
   - Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.
   - Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку 1 H та 2 H за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення 1 H та 2 H на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення 1 H та 2 H на тому ж тексті, в якому вилучено всі пробіли.
   - За допомогою програми CoolPinkProgram оцінити значення $H^{(10)}$, $H^{(20)}$, $H^{(30)}$.
   - Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.
3. Робота виконана 06.10.2021 – 12.10.2021 року в місті Київ. Робота створена на мові програмування Python.
4. Робота над завданням
   - Програми, які підраховують частоти букв та біграм (що перетинаються і не перетинаються). Для визначення кількості букв (пробілу) використав звичайну посимвольну акумуляцію. Для визначення кількості біграм використав контейнер "collections" а саме об'єкт "Counter" в парі з функцією "range(strart, stop, step)" де і зазначав крок дій. В результаті дані використовую для підрахунку їх частотності.

     Складність полягала в недостачі знань роботи в мові програмування. Все було вирішено, прочитавши потрібні інструкції до команд та саму документацію щодо лабораторної.

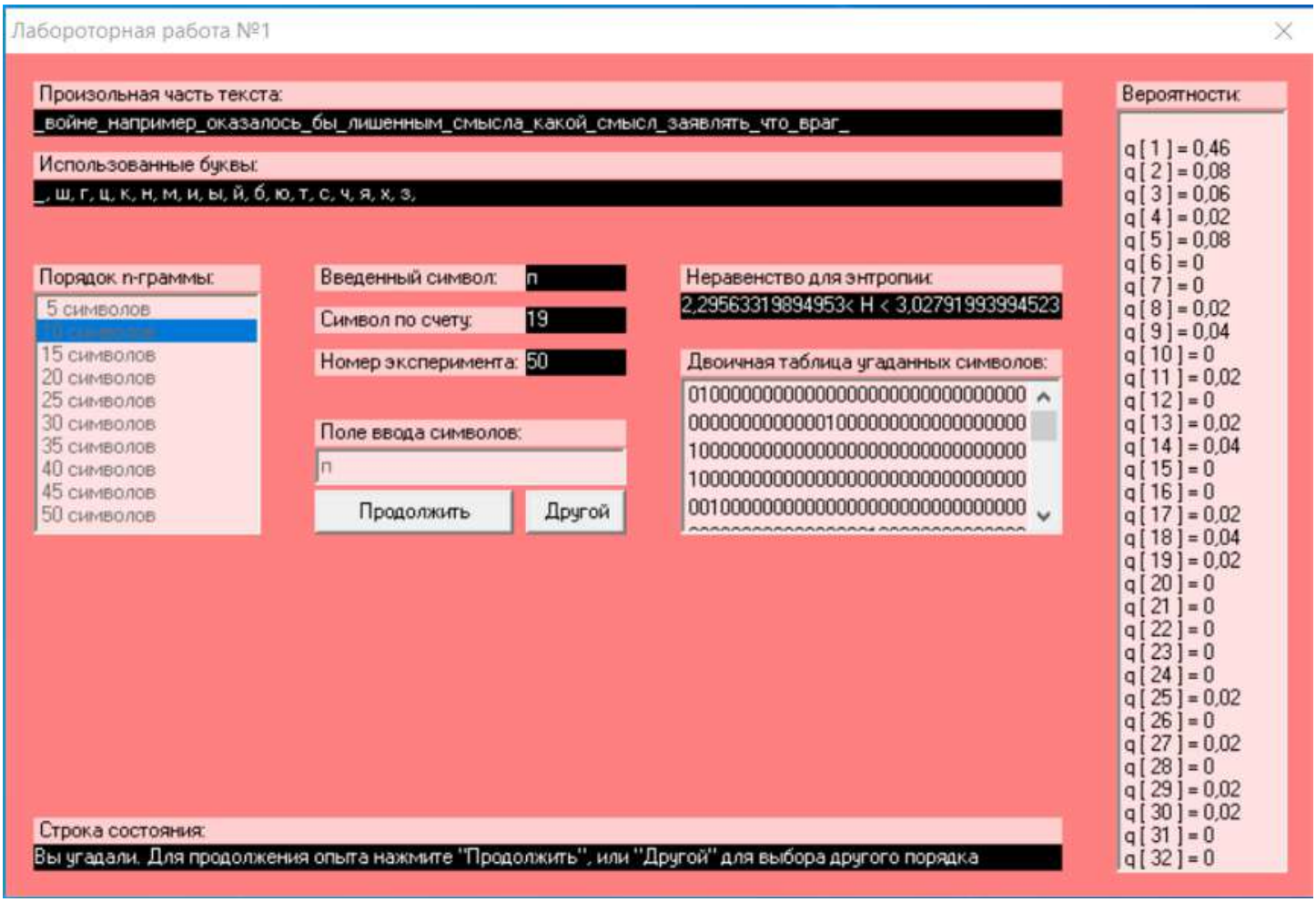
Частота букв без пробілу	**Частота букв з пробілом**
о >>> 0.11275224397101763 time(s)	_ >>> 0.1611100527084524 time(s)
е >>> 0.08463934248945604 time(s)	о >>> 0.0945867240018507 time(s)
а >>> 0.07721855736995782 time(s)	е >>> 0.07100309355977102 time(s)
и >>> 0.06976316643235644 time(s)	а >>> 0.06477787152201327 time(s)
н >>> 0.0640250892181248 time(s)	и >>> 0.058523619011330956 time(s)
т >>> 0.060097328863415164 time(s)	н >>> 0.053710003719529345 time(s)
с >>> 0.050762409430085434 time(s)	т >>> 0.05041504504259315 time(s)
в >>> 0.04583324321401536 time(s)	с >>> 0.042584074971196324 time(s)
л >>> 0.04576186871417757 time(s)	в >>> 0.03844904698400602 time(s)
р >>> 0.042398615767275875 time(s)	л >>> 0.038389171633599145 time(s)
к >>> 0.04061641613496269 time(s)	р >>> 0.03556777254624464 time(s)
д >>> 0.03115388774737753 time(s)	к >>> 0.03407270319063042 time(s)
м >>> 0.03007245593165351 time(s)	д >>> 0.026134683250324326 time(s)
у >>> 0.029680977614361413 time(s)	м >>> 0.0252274809714322 time(s)
п >>> 0.028376770844598248 time(s)	у >>> 0.024899073746473252 time(s)
ь >>> 0.019917811182004974 time(s)	п >>> 0.02380498779812935 time(s)
ы >>> 0.018968314047799285 time(s)	ь >>> 0.01670885157263515 time(s)

я >>> 0.018596301503190225 time(s)	ы >>> 0.015912327971767865 time(s)
ч >>> 0.018496809776143614 time(s)	я >>> 0.015600250387828975 time(s)
б >>> 0.01831296636747053 time(s)	ч >>> 0.015516787778170898 time(s)
г >>> 0.017633827187195847 time(s)	б >>> 0.015362563390759238 time(s)
з >>> 0.01697847950686709 time(s)	г >>> 0.014792840359614983 time(s)
ж >>> 0.010822969611765978 time(s)	з >>> 0.014243075778606356 time(s)
й >>> 0.010444468476262571 time(s)	ж >>> 0.009079280407152383 time(s)
ш >>> 0.009771817886882233 time(s)	й >>> 0.00876175960954014 time(s)
х >>> 0.009546880069211637 time(s)	ш >>> 0.008197479792069237 time(s)
ю >>> 0.006330701849248405 time(s)	х >>> 0.008008781718059675 time(s)
ц >>> 0.003611982264518222 time(s)	ю >>> 0.005310762140634497 time(s)
щ >>> 0.0031664323564399265 time(s)	ц >>> 0.003030055611499696 time(s)
э >>> 0.0027100681302043906 time(s)	щ >>> 0.0026562882725961407 time(s)
ф >>> 0.0015377960419595544 time(s)	э >>> 0.002273448910903664 time(s)
	ф >>> 0.0012900416405846012 time(s)
Ентропія: 4.469409158411811	**Ентропія: 4.386300965219173**
$R = 1 - \frac{4.469}{5} = 0.1062$ 10.62%	$R = 1 - \frac{4.386}{5} = 0.1228$ 12.28%

Біграма з пробілом 1 крок (перші 20)	**Біграма без пробілу 1 крок (перші 20)**
о_ >>> 0.021331993090734906 time(s)	то >>> 0.015845173234937244 time(s)
е_ >>> 0.01837087947977298 time(s)	ко >>> 0.012395398281384841 time(s)
и_ >>> 0.01835273540127444 time(s)	ст >>> 0.01231104641731679 time(s)
_п >>> 0.017079021090676848 time(s)	ов >>> 0.012161808503965619 time(s)
_н >>> 0.016237135848344533 time(s)	но >>> 0.011796283759670725 time(s)
_в >>> 0.015230139491675497 time(s)	по >>> 0.011670837397723365 time(s)
_с >>> 0.015226510675975788 time(s)	на >>> 0.01117986629148111 time(s)
а_ >>> 0.014711218846617217 time(s)	не >>> 0.010842458835208901 time(s)
то >>> 0.01296938731075726 time(s)	ка >>> 0.01019143547406829 time(s)
_и >>> 0.0102350746810271 time(s)	ен >>> 0.009981637248052878 time(s)
ко >>> 0.010193343300480455 time(s)	ни >>> 0.00939982567281426 time(s)
ст >>> 0.010069963566690375 time(s)	го >>> 0.009343591096768892 time(s)
ь_ >>> 0.010048190672492125 time(s)	ак >>> 0.009198678920036596 time(s)
по >>> 0.009785101534263279 time(s)	ос >>> 0.009146770080610102 time(s)
но >>> 0.009652649761223927 time(s)	он >>> 0.009088372636255296 time(s)
_к >>> 0.009612732788527137 time(s)	ро >>> 0.008930483249666377 time(s)
на >>> 0.009307912269751644 time(s)	от >>> 0.008854782858836074 time(s)
_о >>> 0.009258923257805583 time(s)	ал >>> 0.008824502702503952 time(s)
я_ >>> 0.00913010030046594 time(s)	пр >>> 0.008539004085658237 time(s)
не >>> 0.009033936684423672 time(s)	ли >>> 0.008247016863884208 time(s)
Ентропія: 3.9722125027759234	**Ентропія: 4.141235320842601**
$R = 1 - \frac{3.972}{5} = 0.2056$ 20.56%	$R = 1 - \frac{4.141}{5} = 0.1718$ 17.18%

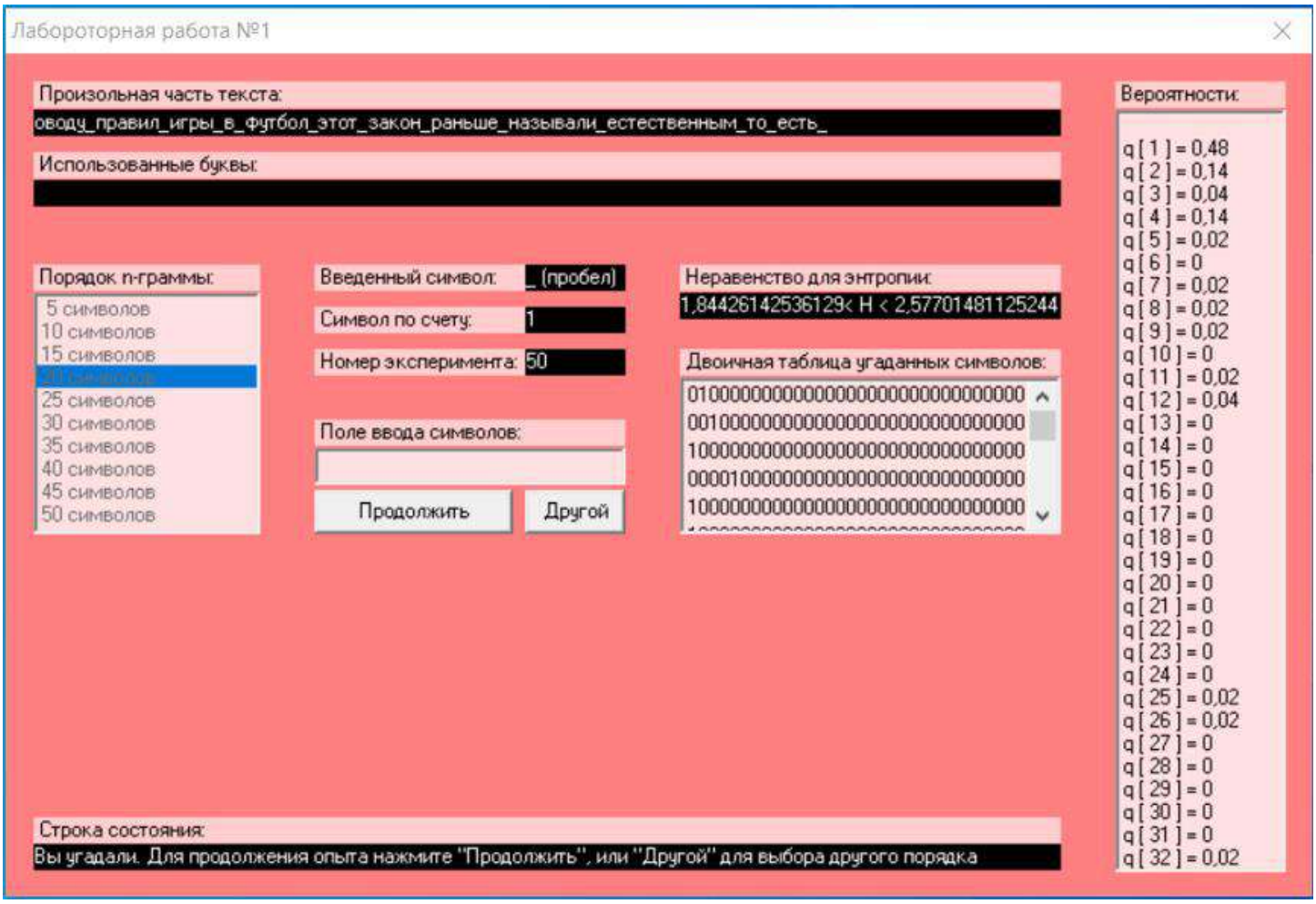
Біграма з пробілом 2 кроки (перші 20)	**Біграма без пробілу 2 кроки (перші 20)**
о_ >>> 0.021333807498584764 time(s)	то >>> 0.015905699145668865 time(s)
и_ >>> 0.018234798891033924 time(s)	ст >>> 0.012821455607223965 time(s)
е_ >>> 0.018209397181135966 time(s)	ов >>> 0.012332648426516709 time(s)
_п >>> 0.017182442338118532 time(s)	ко >>> 0.012233156699470098 time(s)
_н >>> 0.016333299464386802 time(s)	но >>> 0.011982264518222126 time(s)
_в >>> 0.015560361720348947 time(s)	по >>> 0.011489131610251974 time(s)
_с >>> 0.015404322645261492 time(s)	на >>> 0.011099816156591327 time(s)
а_ >>> 0.014489861088935015 time(s)	не >>> 0.010840272520817562 time(s)
то >>> 0.01295124323225872 time(s)	ка >>> 0.01010489888612523 time(s)
_и >>> 0.010334867112769076 time(s)	ен >>> 0.009901589704769115 time(s)
ко >>> 0.010193343300480455 time(s)	ак >>> 0.009529577160160053 time(s)
ь_ >>> 0.009928439754401754 time(s)	ни >>> 0.009473342705742403 time(s)
по >>> 0.009866749887506713 time(s)	го >>> 0.009443062614902131 time(s)
ст >>> 0.009866749887506713 time(s)	ос >>> 0.009066724343030172 time(s)
но >>> 0.009779658310713715 time(s)	от >>> 0.009053747161241485 time(s)
_к >>> 0.009641763314124802 time(s)	ро >>> 0.008962906888720666 time(s)
на >>> 0.009413147925043183 time(s)	он >>> 0.00887206661619985 time(s)
_о >>> 0.009166388457463022 time(s)	ал >>> 0.00871634043473559 time(s)
я_ >>> 0.009075668064970316 time(s)	пр >>> 0.008586568616848709 time(s)
не >>> 0.009035751092273525 time(s)	во >>> 0.007993943981831945 time(s)
Ентропія: 3.9711906476631746 $R = 1 - \frac{3.971}{5} = 0.2058$ 20.58%	**Ентропія: 4.1403186145417195** $R = 1 - \frac{4.140}{5} = 0.172$ 17.2%

- CoolPinkProgram



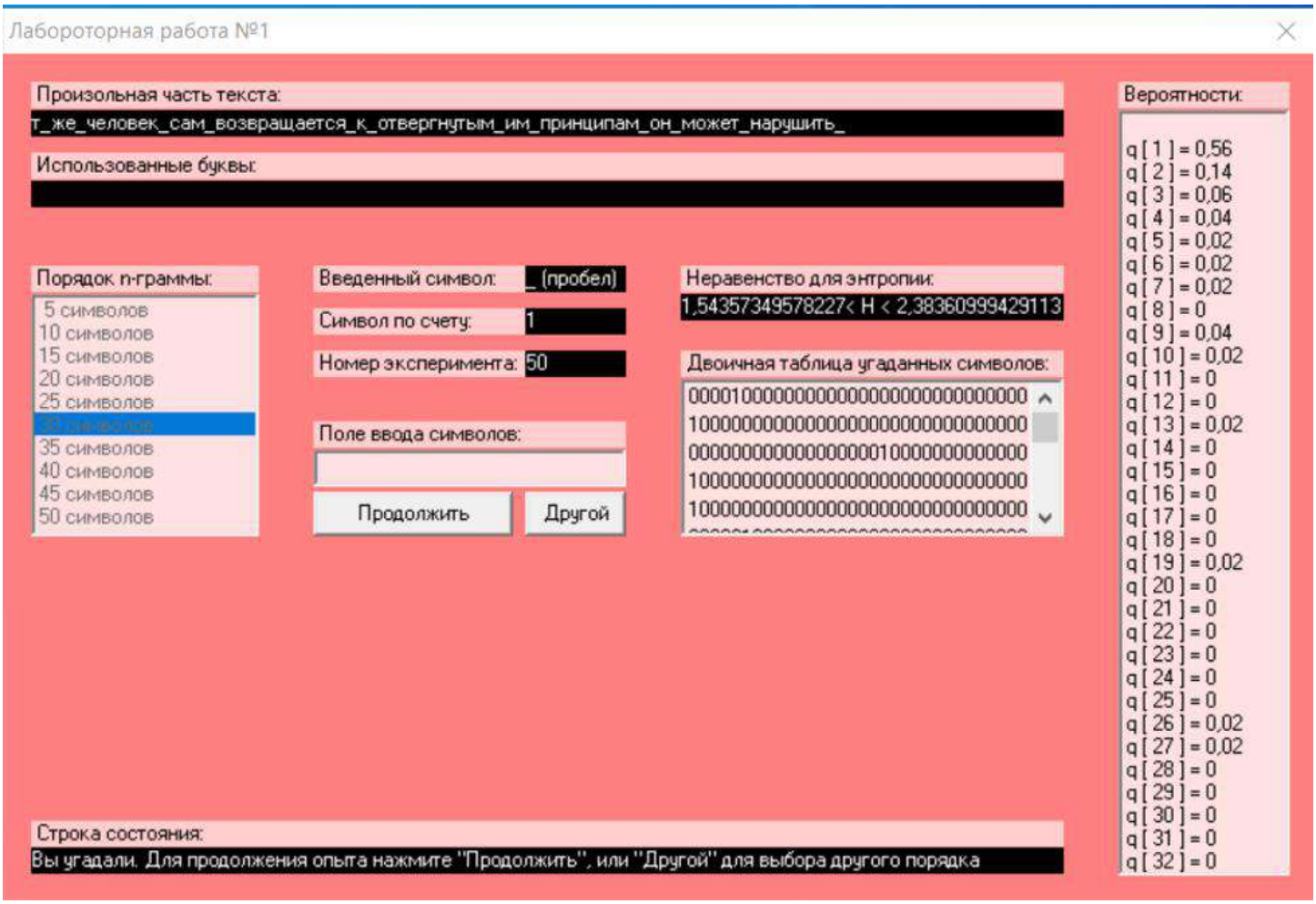
$$2.295 < H < 3.027$$

$$H^{(10)} = \frac{2.295 + 3.027}{2} = 2.661$$



$$1.844 < H < 2.577$$

$$H^{(20)} = \frac{1.844 + 2.577}{2} = 2.2105$$



$$1.543 < H < 2.383$$

$$H^{(30)} = \frac{1.543 + 2.383}{2} = 1.963$$

- Оцінка Надлишковості

$$R_{10} = 1 - \frac{2.661}{5} = 0.4678 \qquad 46.78\%$$

$$R_{20} = 1 - \frac{2.2105}{5} = 0.5579 \qquad 55.79\%$$

$$R_{30} = 1 - \frac{1.963}{5} = 0.6074 \qquad 60.74\%$$

5. Висновки.

Отримав навички роботи з мовою програмування Python. Дізнався як обчислити ентропію, частотність та надлишковість. Розібрався з дослідженням в цілому. Під час дослідження роботи було визначено що найвживаніші букви в російському алфавіті – '_'; 'о'; 'е'; 'а'; щодо менш вживаних – 'щ'; 'ф'; 'э'. Найвживаніші біграми (з пробілом) в російській мові – 'о_'; 'и_'; 'е_'; '_п'; менш вживані (без пробілу) – 'пз'; 'чф'; 'фф'; 'бф'. Також, використавши програму CoolPinkProgram, встановив що середня надлишковвсть російської мови становить 54.43%.